A class of Fejér convergent algorithms, approximate resolvents and the Hybrid Proximal-Extragradient method

B. F. Svaiter *

Abstract

A new framework for analyzing Fejér convergent algorithms is presented. Using this framework we define a very general class of Fejér convergent algorithms and establish its convergence properties. We also introduce a new definition of approximations of resolvents which preserve some useful features of the exact resolvent, and use this concept to present an unifying view of the Forward-Backward splitting method, Tseng's Modified Forward-Backward splitting method and Korpelevich's method. We show that methods based on families of approximate resolvents fall within the aforementioned class of Fejér convergent methods. We prove that such approximate resolvents are the iteration maps of the Hybrid Proximal-Extragradient method.

2010 Mathematics Subject Classification: 47H05, 49J52, 47N10.

Keywords: Fejér convergence, approximate resolvent, Hybrid Proximal-Extragradient

1 Introduction

In this work we introduce a new framework for analysing Fejér convergent algorithms in Hilbert spaces, by means of recursive inclusions and sequences of point-to-set maps. This framework defines a new class of Fejér convergent methods, which is general enough to encompass, for example, the classical Forward-Backward splitting method, Tseng's Modified Forward-Backward splitting method and Korpelevich's method. Using this framework, we prove for good that convergence with summable errors is a generic property of a large class of Fejér convergent algorithms. Therefore, we regard this convergence result (with summable errors) as a rather negative result, in the sense that it is too generic to convey useful information on Fejér convergent methods. For sure, this kind of convergence for particular Fejér convergent algorithms lacks particular value.

Another original contribution of this work is the concept of approximate resolvents of maximal monotone operators. Approximate resolvents retain the relevant features of exact resolvents as iteration maps for finding zeros of maximal monotone operators, their elements are more easily computable and are indeed calculated in, for instance, the classical Forward-Backward splitting method, Tseng's Modified Forward-Backward splitting method and Korpelevich's method. We prove that any algorithm based on approximate resolvents fall within the above mentioned class of Fejér convergent methods, providing an unifying framework for establishing their convergence properties.

We present a new transportation formula for cocoercive operators and use it for establishing that the Forward-Backward splitting method is a particular case of the Hybrid Proximal-Extragradient

*IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil (benar@impa.br) Partially supported by CNPq grants 302962/2011-5, 474944/2010-7, FAPERJ grant E-26/102.940/2011 and by PRONEX-Optimization.

method. The relationship between the Hybrid Proximal-Extragradient method and approximate resolvents is also discussed.

This work is organized as follows. In Section 2, we introduce some basic definitions and results. In Section 3, we define a very general class of Fejér convergent by means of recursive inclusions and sequences of point-to-set maps satisfying two basic properties. In Section 4, we define approximate resolvents, show that they are the iteration maps of the Hybrid Proximal-Extragradient method and prove that methods based on approximate resolvents fall within the aforementioned class of Fejér convergent methods. In Section 5, we show that the Forward-Backward method is based on approximate resolvents, which is to say that it is a particular case of the Hybrid Proximal-Extragradient method. In Section 6, we recall that Tseng's Modified Forward-Backward method is a particular case of the Hybrid Proximal-Extragradient method, which is to say that it is based on approximate resolvents. In Section 7, we recall that Korpelevich's method is a particular case of the Hybrid Proximal-Extragradient method, which is to say that it is based on approximate resolvents. In Section 8, we make some comments.

2 Basic definitions and results

In the first part of this section, we review the concept and properties of Quasi-Fejér convergence, which will be used in our analysis of a class of Fejér convergent methods. In the second part, we establish the notation concerning point-to-set maps, which will be used for defining the aforementioned class.

The last part of this section contains the material needed to define approximate resolvents and the Hybrid Proximal-Extragradient (HPE) method, to prove that methods based on approximate resolvents/HPE method belong to the aforementioned class, and that some well known decomposition methods are based on approximate resolvents/HPE method.

As far as we know, this section contains just one original result, namely, Lemma 2.8, which is a transportation formula for cocoercive operators.

Quasi-Fejér convergence

The concept of Quasi-Fejér convergence was introduce by Ermol′ev [9] in the context of sequences of random variables (see also [10] and its translation [11]). We will use a deterministic version of this notion, considered first in metric spaces by Isuem, Svaiter and Teboulle [12, 13, Definition 4.1], in Euclidean spaces [4, Definition 1], in Hilbert spaces in [2], in reflexive Banach spaces in [1].

All this material is now standard knowledge and is included for the sake of completeness. We do not claim to give *here* any original contribution to this over-studied concept. We will use an arbitrary exponent p just to unify the results related in the references; its particular value seems of little importance as indicated by the next results.

Definition 2.1. *Let X be a metric space and $0 < p < \infty$. A sequence (x_n) in X is p-Quasi-Fejér convergent to $\Omega \subset X$ if, for each $x^* \in \Omega$, there exists a non-negative, summable sequence (ρ_n) such that*

$$d(x^*, x_n)^p \leq d(x^*, x_{n-1})^p + \rho_n \qquad n = 1, 2, \ldots$$

Note that if $\rho_1 = \rho_2 = \cdots = 0$ in the above definition, we retrieve the classical definition of Fejér convergence and the exponent p becomes immaterial. Ermol′lev considered the stochastic case with

$p = 2$ and the deterministic case was considered in [12, 13] with $p = 1$ and in [4, 2] with $p = 2$ in Euclidean and Hilbert spaces respectively. The next proposition summarizes the main properties of Quasi-Fejér convergent sequences in metric spaces.

Proposition 2.2. *Let X be a metric space, $p \in (0, \infty)$ and (x_n) be a sequence in X which is p-Quasi-Fejér convergent to $\Omega \subset X$ then,*

1. *if Ω is non-empty, then (x_n) is bounded;*
2. *for any $x^* \in \Omega$ there exists $\lim_{n\to\infty} d(x^*, x_n) < \infty$;*
3. *if the sequence (x_n) has a cluster point $x^* \in \Omega$, then it converges to such a point.*

Proof. Take $x^* \in \Omega$ and let (ρ_n) be as in Definition 2.1. Then for $n < m$

$$d(x^*, x_m)^p \leq d(x^*, x_n)^p + \sum_{i=n+1}^{m} \rho_i.$$

Hence

$$\limsup_{m\to\infty} d(x^*, x_m)^p \leq d(x^*, x_n)^p + \sum_{i=n+1}^{\infty} \rho_i < \infty,$$

which proves item 1. To prove item 2, note that (ρ_n) is summable and take the $\liminf_{n\to\infty}$ at the right hand side of the first inequality in the above equation. Item 3 follows trivially from item 2. □

Now we recall Opial's Lemma [15], which is useful for analyzing Quasi-Fejér convergence in Hilbert spaces:

Lemma 2.3 (Opial)**.** *If, in a Hilbert space X, the sequence (x_n) is weakly convergent to x^*, then for any $x \neq x^*$*

$$\liminf_{n\to\infty} \|x_n - x\| > \liminf_{n\to\infty} \|x_n - x^*\|.$$

The next result was proved in [16], for the case of a specific sequence generated by an inexact proximal point method with $p = 1$, but the proof presented there is quite general, and we provide it here for the sake of completeness. The idea of using Opial's Lemma seems to be due to H. Brezis. Latter on this result was explicitly proved for Quasi-Fejér convergent sequences in Hilbert and Banach spaces with $p = 2$ in [2, Proposition 1], [1, Lemma 2.8] respectively.

Proposition 2.4. *If, in a Hilbert space X, the sequence (x_n) is p-Quasi-Fejér convergent to $\Omega \subset X$, then it has at most one weak cluster point in Ω.*

Proof. If $x^* \in \Omega$ is a weak cluster point of (x_n), then there exists a subsequence (x_{n_k}) weakly convergent to x^*. Therefore, using item 2 of Proposition 2.2 and Opial's Lemma, we conclude that for any $x' \in \Omega$, $x' \neq x^*$

$$\lim_{n\to\infty} \|x_n - x'\| = \liminf_{k\to\infty} \|x_{n_k} - x'\| > \liminf_{k\to\infty} \|x_n - x^*\| = \lim_{n\to\infty} \|x_n - x^*\|$$

which trivially implies the desired result. □

It is trivial that the specific value of $p \in (0, \infty)$ is immaterial in the above proofs. It would be preposterous to claim that for each p one has a "specific kind" of Quasi-Fejér convergence. We hope to reinforce this point of view with the next remark.

Remark 2.5. *Let X be a metric space, $p \in (0, \infty)$ and (x_n) be a sequence in X which is p-Quasi-Fejér convergent to $\Omega \subset X$ then either,*

1. *$d(x_n, x^*) \to 0$ for some $x^* \in \Omega$;*
2. *(x_n) is q-Quasi-Fejér convergent to Ω for any $q \in (0, \infty)$.*

From now on, p-Quasi-Fejér convergence will be called simply Quasi-Fejér convergence, the exponent p being 1 unless otherwise stated.

Point-to-set operators

Let X, Y be arbitrary sets. A point-to-set map $F : X \rightrightarrows Y$ is a function $F : X \to \wp(Y)$, where $\wp(Y)$ is the power set of Y, that is, the family of all subsets of Y. If $F(x)$ is a singleton for all x, that is, a set with just one element, one says that F is point-to-point. Whenever necessary, we will identify a point-to-point map $F : X \rightrightarrows Y$ with the unique function $f : X \to Y$ such that $F(x) = \{f(x)\}$ for all $x \in X$,

A point-to-set map $F : X \rightrightarrows Y$ is *L-Lipschitz* if X and Y are normed vector spaces and,

$$\emptyset \neq F(x') \subset \{y + u \mid y \in F(x),\ u \in Y,\ \|u\| \leq L\|x - x'\|\}, \qquad \forall x, x' \in X. \tag{1}$$

Note that if F is point-to-point and it is identified with a function, then in the above definition we retrieve the classical notion of a L-Lipschitz continuous function.

Maximal monotone operators and the ε-enlargement

The ε-enlargement of a maximal monotone operators will be used to define approximate resolvents in Section 4. In this section we review the definition of the ε-enlargement and discuss those of its properties which will be used in the analysis and applications of approximate resolvents.

From now on, X is a real Hilbert space. Recall that a point-to-set operator $T : X \rightrightarrows X$ is *monotone* if

$$\langle x - y, u - v\rangle \geq 0 \qquad \forall x, y \in X,\ u \in T(x), v \in T(y),$$

and it is *maximal monotone* if it is monotone and maximal in the family of monotone operators in X with respect to the partial order of the inclusion.

Let $T : X \rightrightarrows X$ be a maximal monotone operator. Recall that the ε-enlargement [5] of T is defined

$$T^{[\varepsilon]}(x) = \{v \mid \langle x - y, v - u\rangle \geq -\varepsilon\}, \qquad x \in X, \varepsilon \geq 0. \tag{2}$$

Now we state some elementary properties of the ε-enlargement which follow trivially from the above definition and the basic properties of maximal monotone operators. Their proofs can be found in [5, 7, 20].

Proposition 2.6. *Let $T : X \rightrightarrows X$ be maximal monotone. Then*

1. *$T = T^{[0]}$;*

2. *if* $0 \le \varepsilon_1 \le \varepsilon_2$*, then* $T^{[\varepsilon_1]}(x) \subset T^{[\varepsilon_2]}(x)$ *for any* $x \in X$*;*

3. $\lambda\left(T^{[\varepsilon]}(x)\right) = (\lambda T)^{[\lambda\varepsilon]}(x)$ *for any* $x \in X$*,* $\varepsilon \ge 0$ *and* $\lambda > 0$*;*

4. *if* $v_k \in T^{[\varepsilon_k]}(x_k)$ *for* $k = 1, 2, \ldots,$ (x_k) *converges weakly to* x*,* (v_k) *converges strongly to* v *and* (ε_k) *converges to* ε*, then* $v \in T^{[\varepsilon]}(x)$*;*

5. *if* $T = \partial f$*, where* f *is a proper closed convex function in* X*, then* $\partial_\varepsilon f(x) \subset T^{[\varepsilon]}(x) = (\partial f)^{[\varepsilon]}(x)$ *for any* $x \in X$*,* $\varepsilon \ge 0$*.*

The ε-enlargements of two operators can be "added" as follows. This fact was proved in [5] in a finite dimensional setting, but its extension to Hilbert and Banach spaces are straightforward.

Proposition 2.7. *It* $T_1, T_2 : X \rightrightarrows X$ *are maximal monotone and* $T_1 + T_2$ *is also maximal monotone then, for any* $\varepsilon_1, \varepsilon_2 \ge 0$ *and* $x \in X$

$$T_1^{[\varepsilon_1]}(x) + T_2^{[\varepsilon_2]}(x) \subset (T_1 + T_2)^{[\varepsilon_1+\varepsilon_2]}(x).$$

Recall that a (maximal) monotone operator $A : X \to X$ is *α-cocoercive* (for $\alpha > 0$) if

$$\langle x - y, Ax - Ay\rangle \ge \alpha\|Ax - Ay\|^2, \qquad \forall x, y \in X.$$

There is an interesting "transportation formula" for cocoercive operators. This result was proved by R.D.C Monteiro and myself.

Lemma 2.8. *If* $A : X \to X$ *is* α*-cocoercive, then for any* $x, z \in X$*,*

$$A(z) \in A^{[\varepsilon]}(x), \qquad \textit{with } \varepsilon = \frac{\|x - z\|^2}{4\alpha}.$$

Proof. Take $y \in X$. Then

$$\begin{aligned}\langle x - y, Az - Ay\rangle &= \langle x - z, Az - Ay\rangle + \langle z - y, Az - Ay\rangle \\ &\ge \langle x - z, Az - Ay\rangle + \alpha\|Az - Ay\|^2 \\ &\ge -\|x - z\|\|Az - Ay\| + \alpha\|Az - Ay\|^2,\end{aligned}$$

where the first inequality follows form the cocoercivity of A and the second one from Cauchy-Schwarz inequality. To end the proof, note that

$$-\|x - z\|\|Az - Ay\| + \alpha\|Az - Ay\|^2 \ge \inf_{t\in\mathbb{R}} \alpha t^2 - \|x - z\|t$$

and compute the value of the left hand-side of this inequality. □

The usefulness of the σ-approximate resolvent (to be defined in Section 4) follows from the next elementary result, essentially proved in [17, Lemma 2.3, Corollary 4.2].

Lemma 2.9. *Suppose that* $T : X \rightrightarrows X$ *is maximal monotone,* $x \in X$*,* $\lambda > 0$ *and* $\sigma \ge 0$*. If*

$$\begin{cases} v \in T^{[\varepsilon]}(y), \\ \|\lambda v + y - x\|^2 + 2\lambda\varepsilon \le \sigma^2\|y - x\|^2, \end{cases} \qquad \textit{and} \qquad z = x - \lambda v,$$

then $\|\lambda v\| \le (1 + \sigma)\|y - x\|$*,* $\|z - y\| \le \sigma\|y - x\|$ *and for any* $x^* \in T^{-1}(0)$

$$\begin{aligned}\|x^* - x\|^2 &\ge \|x^* - z\|^2 + \|y - x\|^2 - \left[\|\lambda v + y - x\|^2 + 2\varepsilon\right] \\ &\ge \|x^* - z\|^2 + (1 - \sigma^2)\|y - x\|^2.\end{aligned}$$

Proof. Since $\varepsilon \geq 0$ we have $\|\lambda v + y - x\| \leq \sigma \|y - x\|$. The two first inequalities of the lemma follows trivially from this inequality, triangle inequality and the definition of z.

To prove the third inequality of the lemma, take $x^* \in T^{-1}(0)$. Direct combination of the algebraic identities

$$\begin{aligned} \|x^* - x\|^2 &= \|x^* - z\|^2 + 2\langle x^* - y, z - x\rangle + 2\langle y - z, z - x\rangle + \|z - x\|^2 \\ &= \|x^* - z\|^2 + 2\langle x^* - y, z - x\rangle + \|y - x\|^2 - \|y - z\|^2 \end{aligned}$$

with the definition of z yields

$$\|x^* - x\|^2 = \|x^* - z\|^2 + 2\lambda \langle x^* - y, -v\rangle + \|y - x\|^2 - \|\lambda v + y - x\|^2.$$

Using the inclusions $0 \in T(x^*)$, $v \in T^{[\varepsilon]}(y)$ and the definition in (2), we conclude that $\langle x^* - y, 0 - v\rangle \geq -\varepsilon$. To end the proof, of the third inequality, combine this inequality with the above equations.

The last inequality follows trivially from the third one and the assumptions of the lemma. $\square$

3 A class of Fejér convergent methods

Let X be a Hilbert space and $\Omega \subset X$. We are concerned with iterative methods for solving problem

$$x \in \Omega. \tag{3}$$

These methods, in their exact or inexact form, generate sequences (x_n) by means of the *recursive inclusions*

$$x_n \in F_n(x_{n-1}) \text{ or } x_n \in F_n(x_{n-1}) + r_n, \qquad n = 1, 2, \ldots,$$

respectively, where $F_1 : X \rightrightarrows X, F_2 : X \rightrightarrows X, \ldots$ are point-to-set maps, $r_1, r_2 \ldots$ are errors and $x_0 \in X$ is a starting point. The basic elements here are the set Ω and the family of point-to-set maps (F_n), which we will call the family of *iteration-maps*.

We will consider two properties of a general family of point-to-set maps $(F_n : X \rightrightarrows X)_{n\in\mathbb{N}}$ with respect to $\Omega \subset X$:

P1: if $\hat{x} \in F_n(x)$ and $x^* \in \Omega$ then

$$\|x^* - \hat{x}\| \leq \|x^* - x\|;$$

P2: if $(z_k)_{k\in\mathbb{N}}$ converges weakly to $\bar{z}$, $\hat{z}_k \in F_{n_k}(z_k)$ for $n_1 < n_2 < \cdots$ and for some $x^* \in \Omega$

$$\lim_{k\to\infty} \|x^* - z_k\| - \|x^* - \hat{z}_k\| = 0,$$

then $\bar{z} \in \Omega$.

Property **P1** ensures that points in the image of $F_n(x)$ are closer (or no more distant) to Ω than x. Regarding property **P2**, note that (using property **P1**) we have

$$\|x^* - z_k\| - \|x^* - \hat{z}_k\| \geq 0.$$

The left hand-side of the above inequality measures the progress of $\hat{z}_k$ toward the solution x^*, as compared to z_k. Hence, property **P2** ensures that if the progress becomes "negligible", then the weak limit point of (z_k) belongs to Ω.

Theorem 3.1. *Suppose that $\Omega \subset X$ is non-empty and $(F_n : X \rightrightarrows X)$ is a sequence of point-to-set maps which satisfies conditions* **P1**, **P2** *with respect to Ω.*

If

$$x_n \in F_n(x_{n-1}) + r_n, \qquad \sum \|r_n\| < \infty$$

then (x_n) is Quasi-Fejér convergent to Ω, it converges weakly to some $\bar{x} \in \Omega$ and for any $w \in \Omega$ there exists $\lim_{n\to\infty} \|x^* - x_n\|$.

Moreover, if $r_n = 0$ for all n, then (x_n) is Fejér convergent to Ω.

Proof. To simplify the proof, define

$$\hat{x}_n = x_n - r_n.$$

Take an arbitrary $x^* \in \Omega$. Since $\hat{x}_n \in F_n(x_{n-1})$, $\|x^* - \hat{x}_n\| \leq \|x^* - x_{n-1}\|$,

$$\|x^* - x_n\| \leq \|x^* - \hat{x}_n\| + \|r_n\| \leq \|x^* - x_{n-1}\| + \|r_n\|$$

and $(x_n)_{n\in\mathbb{N}}$ is Quasi-Fejér convergent to Ω. Therefore, by Proposition 2.2, this sequence is bounded and there exists $\lim_{n\to\infty} \|x^* - x_n\| < \infty$. Using this fact, the above equation and the assumption of (r_n) being summable we conclude that

$$\lim_{n\to\infty} \|x^* - x_{n-1}\| - \|x^* - \hat{x}_n\| = 0.$$

Since $(x_n)_{n\in\mathbb{N}}$ is bounded it has a weak cluster point, say $\bar{x}$ and there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ which converges weakly to $\bar{x}$. The above equation shows that, in particular

$$\lim_{k\to\infty} \|x^* - x_{n_k-1}\| - \|x^* - \hat{x}_{n_k}\| = 0$$

Therefore, using **P2**, the two above equations and the inclusion $\hat{x}_{n_k} \in F_{n_k}(x_{n_k-1})$, we conclude that $\bar{x} \in \Omega$. Hence, all weak cluster points of (x_n) belong to Ω. To end the proof, use Proposition 2.4 □

Note that properties **P1**, **P2** are "inherited" by specializations. We state formally this result and the proof, being quite trivial, is be omitted

Proposition 3.2. *If $(F_n : X \rightrightarrows X)$ is a sequence of point-to-set maps which satisfies conditions* **P1**, **P2** *with respect to $\Omega \subset X$ and $(G_n : X \rightrightarrows X)$ is a sequence of point-to-set maps such that, for any $x \in X$*

$$G_n(x) \subset F_n(x), \qquad n = 1, 2, \ldots,$$

then $(G_n : X \rightrightarrows X)$ also satisfies conditions **P1**, **P2** *with respect to Ω.*

What about compositions? Suppose that (F_n) is a sequence satisfying **P1**, **P2** with respect to $\Omega \subset X$, and that

$$F_n = G_n \circ H_n$$

where $G_n : X \rightrightarrows Y$, $H_n : Y \rightrightarrows X$ and all G_n's are L-Lipschitz continuous (Y is Hilbert). One may consider sequences

$$y_n \in H_n(x_{n-1}) + u_n, \qquad x_n \in G_n(y_n) + r_n, \tag{4}$$

where (u_n) and (r_n) are summable. Since $x_n - r_n \in G_n(y_n)$, using also (1), we conclude that there exists $\hat{x}_n \in G_n(y_n - u_n) \subset G_n \circ H_n(x_{n-1})$

$$\|\hat{x}_n - (x_n - r_n)\| \leq L \, \|u_n\|.$$

Therefore, defining $s_n = x_n - \hat{x}_n$ we conclude that

$$x_n \in F_n(x_{n-1}) + s_n, \qquad \sum \|s_n\| \leq \sum L\|u_n\| + \|r_n\| < \infty.$$

Therefore, if $\Omega \neq \emptyset$, a sequence (x_n) generated as in (4) converges weakly to some point $x^* \in \Omega$.

On may also consider compositions of $m+1$ maps

$$F = G_{1,n} \circ G_{2,n} \cdots \circ G_{m,n} \circ H_n$$

adding summable errors in each stage, assuming each $G_{i,n} : Y_i \rightrightarrows Y_{i-1}$ to be L-Lipschitz continuous, $H_n : X \rightrightarrows Y_m$, $Y_0 = X$ etc.

4 Approximate resolvents and the Hybrid Proximal-Extragradient Method

In this section, first we define σ-approximate resolvents, analyze some of their properties and study conditions under which sequences of σ-approximate resolvents satisfy properties **P1**, **P2**. After that, we recall the definition of the Hybrid Proximal-Extragradient method and show that σ-approximate resolvents are the iteration maps of such method. At the end of the section we discuss the incorporation of summable errors to sequences of σ-approximate resolvents and to the Hybrid Proximal-Extragradient method.

Recall that the *resolvent* of a maximal monotone operator $T : X \rightrightarrows X$ is defined as

$$J_T(x) = (I + T)^{-1}(x), \qquad x \in X. \tag{5}$$

We shall consider approximations of the resolvent in the following sense.

Definition 4.1. *The* σ-approximate resolvent *of a maximal monotone operator* $T : X \rightrightarrows X$ *is the point-to-set operator* $J_{T,\sigma} : X \rightrightarrows X$

$$J_{T,\sigma}(x) = \left\{ x - v \;\middle|\; \begin{array}{l} \exists \varepsilon \geq 0, y \in X, \\ v \in T^{[\varepsilon]}(y) \\ \|v + y - x\|^2 + 2\varepsilon \leq \sigma^2 \|y - x\|^2 \end{array} \right\}$$

where $\sigma \geq 0$.

First, we analyze some elementary properties of approximate resolvents and find a convenient expression for $J_{\lambda T,\sigma}$. In particular, we show that the σ-approximate resolvent is indeed and extension (in the sense of point-to-set maps) of the classical resolvent.

Proposition 4.2. *Let* $T : X \rightrightarrows X$ *be maximal monotone. Then, for any* $x \in X$,

1. $J_{T,\sigma=0}(x) = \{J_T(x)\}$;

2. *if* $0 \leq \sigma_1 \leq \sigma_2$ *then* $J_{T,\sigma_1}(x) \subset J_{T,\sigma_2}(x)$*;*

3. *for any* $\lambda > 0$ *and* $\sigma \geq 0$*,*

$$J_{\lambda T,\sigma}(x) = \left\{ x - \lambda v \;\middle|\; \begin{array}{l} \exists \varepsilon \geq 0, y \in X, \\ v \in T^{[\varepsilon]}(y) \\ \|\lambda v + y - x\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|y - x\|^2 \end{array} \right\}$$

Proof. Items 1, 2 and 3 follow trivially from Definition 4.1 and Proposition 2.6, items 1, 2 and 3. ☐

Note that in view of item 1 of the above proposition, if point-to-set operators which are point-to-point are identified with functions, we have

$$J_{T,0} = J_T\,.$$

In view of item 3,

$$J_{\lambda T,\sigma} = \left\{ z \in X; \;\middle|\; \begin{array}{l} \exists \varepsilon \geq 0, y \in X, \\ \dfrac{x - z}{\lambda} \in T^{[\varepsilon]}(y) \\ \|y - z\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|y - x\|^2 \end{array} \right\}$$

The next theorem is the main result of this section and states that, in some sense, approximate resolvents are "almost as good" as resolvents for finding zeros of maximal monotone operators, that is, for solving problem (3) with $\Omega = \{x \mid 0 \in T(x)\} = T^{-1}(0)$.

Theorem 4.3. *Suppose that* $T : X \rightrightarrows X$ *is maximal monotone,* $\sigma \in [0, 1)$*,* $\underline{\lambda} > 0$ *and* (λ_k) *is a sequence in* $[\underline{\lambda}, \infty)$*. Then, the sequence of point-to-set maps*

$$(J_{\lambda_k T,\sigma})_{k \in \mathbb{N}}$$

*satisfies properties **P1**, **P2** with respect to* $\Omega = \{x \in X \mid 0 \in T(x)\} = T^{-1}(0)$*.*

Proof. Suppose that $\hat{x} \in J_{\lambda_k T,\sigma}(x)$. This means that there exists $y, v \in X$, $\varepsilon \geq 0$ such that

$$\hat{x} = x - \lambda_k v, \quad v \in T^{[\varepsilon]}(x), \quad \|\lambda_k v + y - x\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|y - x\|^2.$$

Therefore, using Lemma 2.9 we conclude that for any $x^* \in= T^{-1}(0)$,

$$\|x^* - x\|^2 \geq \|x^* - \hat{x}\|^2 + (1 - \sigma^2)\|y - x\|^2 \geq \|x^* - \hat{x}\|^2$$

which proves that the family $(J_{\lambda_k T,\sigma})$ satisfies **P1**.

Now we prove **P2**. Suppose that (z_k) converges weakly to $\bar{z}$, $\hat{z}_k \in J_{\lambda_{n_k} T,\sigma}(z_k)$, $0 \in T(x^*)$ and

$$\lim_{k \to \infty} \|x^* - z_k\| - \|x^* - \hat{z}_k\| = 0. \tag{6}$$

To simplify the proof, let $\mu_k = \lambda_{n_k} \geq \underline{\lambda}$. For each k there exists $v_k, y_k \in X$, $\varepsilon_k \geq 0$ such that

$$\hat{z}_k = z_k - \mu_k v, \quad v_k \in T^{\varepsilon_k}(z_k), \quad \|\mu_k v_k + y_k - z_k\|^2 + 2\mu_k\varepsilon \leq \sigma^2 \|y_k - z_k\|^2. \tag{7}$$

Using again Lemma 2.9, we conclude that

$$\|x^* - z_k\|^2 \geq \|x^* - \hat{z}_k\|^2 + (1 - \sigma^2)\|y_k - z_k\|^2.$$

Therefore

$$(1-\sigma^2)\|y_k - z_k\|^2 \leq \|x^* - z_k\|^2 - \|x^* - \hat{z}_k\|^2$$
$$= (\|x^* - z_k\| - \|x^* - \hat{z}_k\|)(\|x^* - z_k\| + \|x^* - \hat{z}_k\|).$$

Since (z^k) is weakly convergent, it is also bounded. Taking this fact into account and using the above equation and (6) we conclude that

$$\lim_{k\to\infty} \|y_k - z_k\| = 0.$$

So, (y_k) also converges weakly to $\bar{z}$. Since $\varepsilon_k \geq 0$, using the last relation in (7) we conclude that

$$\mu_k\varepsilon_k \leq \frac{\sigma^2}{2}\|y_k - z_k\|^2, \qquad \|\mu_k v_k\| \leq (1+\sigma)\|y_k - z_k\|.$$

Therefore, since (μ_k) is bounded away from 0,

$$\lim_{k\to\infty} \varepsilon_k = 0, \quad \lim_{k\to\infty} v_k = 0$$

and $0 \in T^{[0]}(\bar{z}) = T(\bar{z})$. □

The Hybrid Proximal-Extragradient/Projection methods were introduced in [18, 17, 19]. These methods are variants of the Proximal Point method which use relative error tolerances for accepting inexact solutions of the proximal sub-problems. Here we are concerned with the variant introduced in [17], which will be called, from now on, the Hybrid Proximal-Extragradient (HPE) method. It solves iteratively the problem

$$0 \in T(x), \tag{8}$$

where $T : X \rightrightarrows X$ is maximal monotone. This method proceeds as follows.

Algorithm: (projection free) HPE method [17]:
Choose $x_0 \in X$, $\sigma \in [0,1)$, $\underline{\lambda} > 0$ and for $k = 1, 2, \ldots$
a) Choose $\lambda_k \geq \underline{\lambda}$ and find/compute $v_k, y_k \in X$, $\varepsilon \geq 0$ such that

$$v_k \in T^{\varepsilon_k}(y_k), \qquad \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k\varepsilon_k \leq \sigma^2\|y_k - x_{k-1}\|^2$$

b) Set $x_k = x_{k-1} - \lambda_k v_k$

To generate iteratively sequences by means of approximate resolvents is equivalent to apply the HPE method in the following sense.

Proposition 4.4. *Let $T : X \rightrightarrows X$ be maximal monotone, $\sigma \geq 0$, $\underline{\lambda} > 0$ and (λ_k) be sequence in $[\underline{\lambda}, \infty)$.*

A sequence (x_k) satisfies the recurrent inclusion

$$x_k \in J_{\lambda_k T, \sigma}(x_{k-1}), \qquad k = 1, 2, \ldots$$

if and only if there exists sequences (y_k), (v_k), (ε_k) which, together with the sequences (x_k), (λ_k) satisfy steps a) and b) of the HPE method.

Proof. Use Definition 4.1 and Proposition 4.2 item 3. □

Convergence of the HPE method perturbed by a summable sequence of errors was proved directly in [6]. Here we see that it can be easily and effortlessly derived as a particular case of a generic convergence result, combining Proposition 4.4 with Theorem 4.3.

Corollary 4.5. *If $T : X \rightrightarrows X$ is maximal monotone, $T^{-1}(0) \neq \emptyset$, $\underline{\lambda} > 0$, $\sigma \in [0, 1)$, for $k = 1, 2, \ldots$*

$$\begin{aligned}
&\lambda_k \geq \underline{\lambda}\\
&v_k \in T^{[\varepsilon_k]}(\tilde{x}_k),\ \|\lambda_k v_k + \tilde{x}_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma \|\tilde{x}_k - x_{k-1}\|^2\\
&x_k = x_{k-1} - \lambda_k v_k + r_k
\end{aligned}$$

and $\sum \|r_k\| < \infty$, then (x_k) (and $(\tilde{x}_k)$) converges weakly to a point $\bar{x} \in T^{-1}(0)$.

Corollary 4.6. *If $T : X \rightrightarrows X$ is maximal monotone, $T^{-1}(0) \neq \emptyset$, $\overline{\lambda} \geq \underline{\lambda} > 0$, $\sigma \in [0, 1)$, for $k = 1, 2, \ldots$*

$$\begin{aligned}
&\overline{\lambda} \geq \lambda_k \geq \underline{\lambda}\\
&v_k \in T^{[\varepsilon_k]}(\tilde{x}_k),\ \|\lambda_k v_k + \tilde{x}_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma \|\tilde{x}_k - x_{k-1}\|^2\\
&x_k = x_{k-1} - \lambda_k (v_k + r_k)
\end{aligned}$$

and $\sum \|r_k\| < \infty$, then (x_k) (and $(\tilde{x}_k)$) converges weakly to a point $\bar{x} \in T^{-1}(0)$.

5 The Forward-Backward splitting method

We will prove in this section that the iteration maps of the Forward-Backward splitting method are specializations or selections of σ-approximate resolvents and the sequence of iteration maps satisfies properties **P1**, **P2**. Equivalently, the Forward-Backward splitting method is a particular instance of the HPE method. Observe that, as a consequence, sequences generated by the inexact Forward-Backward splitting methods with summable errors still converge weakly to solutions of the inclusion problem, if any. This convergence result was previously obtained in [8] by a detailed analysis of the Forward-Backward splitting method. Here we see that it can be easily and effortlessly derived as a particular case of a generic convergence result.

The Forward-Backward Splitting method solves the inclusion problem

$$0 \in (A + B)x$$

where

f1) $A : X \to X$ is α-cocoercive, $\alpha > 0$;
f2) $B : X \rightrightarrows X$ is maximal monotone.

This method proceeds as follows:

FORWARD-BACKWARD SPLITTING METHOD
0) Initialization: Choose $0 < \underline{\lambda} \leq \bar{\lambda} < 2\alpha$ and $x_0 \in X$;
1) for $k = 1, 2, \ldots$
a) choose $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$ and define

$$\begin{aligned}
x_k &= (I + \lambda_k B)^{-1}(x_k - \lambda_k A(x_{k-1}))\\
&= J_{\lambda_k B} \circ (I - \lambda_k A)\ (x_{k-1}). \qquad (9)
\end{aligned}$$

Note that the generic iteration map of the Forward-Backward method is

$$J_{\lambda B} \circ (I - \lambda A) \tag{10}$$

with $\lambda = \lambda_k$ in the k-th iteration.

Lemma 5.1. *If A, B satisfy* **f1** *and* **f2** *then, for any $\lambda > 0$ and $x \in X$,*

$$J_{\lambda B} \circ (I - \lambda A)(x) \in J_{\lambda(A+B),\sigma}(x),$$

with $\sigma = \sqrt{\lambda/(2\alpha)}$.

Proof. Take $x \in X$ and let $z = J_{\lambda B} \circ (I - \lambda A)(x)$. This means that

$$b := \lambda^{-1}(x - \lambda A(x) - z) \in B(z).$$

Define $\varepsilon = \|x - z\|^2/(4\alpha)$, $v = A(x) + b$. Using Lemma 2.8 we conclude that $A(x) \in A^{[\varepsilon]}(z)$. Therefore, combining this result with these two definitions, the above equation, Proposition 2.7 and Proposition 2.6 item 1, we conclude that

$$\begin{aligned} &v \in (A^{[\varepsilon]} + b)(z) \subset (A + B)^{[\varepsilon]}(z), \quad \|\lambda v + z - x\|^2 + 2\lambda\varepsilon = \sigma^2\|z - x\|^2 \\ &z = x - \lambda v, \end{aligned}$$

which, together with Proposition 4.2 item 3, proves the lemma. □

Corollary 5.2. *Let A, B be as in* **f1**, **f2** *and $\underline{\lambda}$, $\bar{\lambda}$ and (λ_k), (x_k) be as in the Forward Backward method. Define*

$$\sigma = \sqrt{\bar{\lambda}/(2\alpha)}.$$

Then $0 < \sigma < 1$ and for any $x \in X$

$$J_{\lambda_k B} \circ (I - \lambda_k A)(x) \in J_{\lambda_k(A+B),\sigma}(x), \qquad k = 1, 2, \ldots \tag{11}$$

In particular

$$x_k = J_{\lambda_k B} \circ (I - \lambda_k A)(x_k) \in J_{\lambda_k(A+B),\sigma}(x_k), \qquad k = 1, 2, \ldots \tag{12}$$

and the sequence of maps $(J_{\lambda_k B} \circ (I - \lambda_k A))$ satisfies properties **P1**, **P2** *with respect to $(A+B)^{-1}(0)$.*

Proof. The bounds for σ follow trivially from its definition and the choices for $\underline{\lambda}$, $\bar{\lambda}$ in the Forward-Backward method.

Define $\sigma_k = \sqrt{\lambda_k/(2\alpha)}$ for $k = 1, 2, \ldots$ Since $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$, $0 < \sigma_k \leq \sigma$ for all k. Therefore, using also Lemma 5.1 and Proposition 4.2 item 2, we conclude that for any $x \in X$

$$J_{\lambda_k B} \circ (I - \lambda_k A)(x) \in J_{\lambda_k(A+B),\sigma_k}(x) \subset J_{\lambda_k(A+B),\sigma}(x), \quad k = 1, 2, \ldots$$

The equality in (12) follows trivially from the definition of the Forward-Backward method, while the inclusion follows from the above equation. To end the proof, note that $0 < \underline{\lambda} < \lambda_k$ for all k, and use Theorem 4.3, Proposition 3.2 and the above equation. □

Proposition 5.3. *Let* (λ_k), (x_k) *be sequences generated by the Forward-Backward Splitting method. Define*

$$\sigma = \sqrt{\frac{\overline{\lambda}}{2\alpha}}, \quad v_k = \lambda_k^{-1}(x_{k-1} - x_k), \quad \varepsilon_k = \frac{\|x_k - x_{k-1}\|^2}{4\alpha}, \ k = 1, 2, \dots$$

Then $0 < \sigma < 1$ *and for* $k = 1, 2, \dots$

$$v^k \in (B + A)^{[\varepsilon_k]}(x_k), \quad \|\lambda_k v_k + x_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma \|x_k - x_{k-1}\|^2$$
$$x_k = x_{k-1} - \lambda_k v_k.$$

In particular, the Forward-Backward splitting method above defined is a particular case of the HPE method with $\sigma \in (0, 1)$.

Proof. See the proofs of Lemma 5.1 and Corollary 5.2. □

6 Tseng's Modified Forward-Backward splitting method

In [17] it was proved that Tseng's Modified Forward-Backward splitting method [21] is a particular case of the HPE method. Here we will cast this result in the framework of approximate resolvents, prove that the iteration maps of the Tseng's Modified Forward-Backward splitting method are specializations or selections of σ-approximate resolvents and that the sequence of its iteration maps satisfies properties **P1**, **P2**. Observe that, as a consequence, sequences generated by inexact Tseng's Modified Forward-Backward splitting method with summable errors still converge weakly to solutions of the inclusion problem, if any. This result follows also from the fact that Tseng's method is a particular case of the HPE (as proved in [17]) and that the HPE with summable errors converges (as proved in [6]). Convergence of Tseng's Modified Forward-Backward splitting method with summable errors was obtained in [3] by a detailed analysis of the Tseng's Modified Forward-Backward splitting method. Here we see that this result can be easily and effortlessly derived as a particular case of a generic convergence result.

In this section we consider the inclusion problem

$$0 \in (A + B)\, x$$

where

t1) $A : X \to X$ is monotone and L-Lipschitz continuous ($L > 0$);
t2) $B : X \rightrightarrows X$ is maximal monotone.

The exact Tseng's Modified Forward-Backward Splitting method (without the auxiliary projection step) proceeds as follows:

Tseng's Modified Forward-Backward method [21]
Choose $0 < \underline{\lambda} \leq \bar{\lambda} < 1/L$ and $x_0 \in X$;
for $k = 1, 2, \dots$
a) choose $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$ and compute

$$y_k = (I + \lambda_k B)^{-1}(x_{k-1} - \lambda_k A(x_{k-1})), \qquad x_k = y_{k-1} - \lambda_k (A(y_k) - A(x_{k-1})).$$

In order to cast this method in the formalism of Section 3, define for $\lambda > 0$

$$H_\lambda : X \to X \times X, \qquad H_\lambda(x) = (x, J_{\lambda B}(x - \lambda A(x))) \tag{13}$$
$$G_\lambda : X \times X \to X, \qquad G_\lambda(x, y) = y - \lambda(A(y) - A(x))) \tag{14}$$

Note that the second component of the (generic) operator H_λ is $J_{\lambda B} \circ (I - \lambda A)$ which is the generic iteration map of the Forward-Backward method in (10). Trivially,

$$x_k = G_{\lambda_k} \circ H_{\lambda_k}\ (x_{k-1}). \tag{15}$$

The next two result were essentially proved in [17], in the context of the Hybrid Proximal-Extragradient Method. We will state and prove it in the context of σ-approximate resolvents.

Lemma 6.1. *If A, B satisfy assumptions **t1**, **t2**, then, for any $\lambda > 0$ and $x \in X$*

$$G_\lambda \circ H_\lambda(x) \in J_{A+B,\sigma}\ (x)$$

for $\sigma = \lambda L$.

Proof. Take $x \in X$ and let

$$y = J_{\lambda B}(x - \lambda A(x)), \qquad z = y + \lambda(A(y) - A(x)).$$

Note that $z = G_\lambda \circ H_\lambda(x)$. Using the definition of y we have

$$a := \lambda^{-1}(x - \lambda A(x) - y) \in B(y).$$

Therefore,

$$v := a + A(y) \in (A + B)(y), \qquad \begin{aligned} \|\lambda v + x - y\|^2 &= \|\lambda(A(y) - A(x))\|^2 \\ &\leq (\lambda L)^2 \|y - x\|^2, \end{aligned}$$

where the inequality follows from assumption **t2)**. To end the proof, note that $z = x - \lambda v$. $\square$

Corollary 6.2. *Let A, B be as in **t1**, **t2** and $0 < \underline{\lambda} < \bar{\lambda} < 2\alpha$ and (λ_k), (x_k) be as in Tseng's Modified Forward-Backward method. Define*

$$\sigma = \bar{\lambda} L.$$

Then $0 < \sigma < 1$ and for any $x \in X$

$$G_{\lambda_k} \circ H_{\lambda_k}(x) \in J_{\lambda_k(A+B),\sigma}\ (x), \qquad k = 1, 2, \ldots \tag{16}$$

In particular

$$x_k = G_{\lambda_k} \circ H_{\lambda_k}(x_{k-1}) \in J_{\lambda_k(A+B),\sigma}\ (x_{k-1}), \qquad k = 1, 2, \ldots \tag{17}$$

*and the sequence of maps $(G_{\lambda_k} \circ H_{\lambda_k})$ satisfies properties **P1**, **P2** with respect to $(A + B)^{-1}(0)$.*

Proof. The bounds for σ follow trivially from its definition and the choices for $\underline{\lambda}$ and $\bar{\lambda}$ in Tseng's Forward-Backward method.

Define $\sigma_k = \lambda_k L$ for $k = 1, 2, \ldots$ Since $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$ we have $0 < \sigma_k \leq \sigma$ for all k. Therefore, using also Lemma 6.1 and Proposition 4.2 item 2, we conclude that for any $x \in X$

$$J_{\lambda_k B} \circ (I - \lambda_k A)(x) \in J_{\lambda_k (A+B),\sigma_k}(x) \subset J_{\lambda_k (A+B),\sigma}(x), \quad k = 1, 2, \ldots$$

The equality in (17) follows trivially from the definition of the Tseng's Modified Forward-Backward method, while the inclusion follows from the above equation. To end the proof, note that $0 < \underline{\lambda} < \lambda_k$ for all k, and use Theorem 4.3, Proposition 3.2 and the above equation. $\square$

Note that for $0 < \lambda \leq \bar{\lambda}$, the maps H_λ, G_λ are Lipschitz continuous with constant

$$2 + \bar{\lambda}L, \qquad 1 + 2\bar{\lambda}L,$$

respectively. Hence, this method can be perturbed by summable sequences of errors in the evaluations of the resolvents $J_{\lambda_k B}$ and/or in the evaluation of $A(x_k)$, $A(y_k)$ etc, and will still converge weakly to a solution, if any exists.

7 Korpelevich's method

In [14] it was proved that Korpelevich's method, with fixed stepsize, is a particular case of the HPE method. The extension of this result for variable stepsizes is trivial, and here we will analyze such an extension in the framework of approximate resolvents. Observe that, as a consequence, sequences generated by inexact Korpelevich's method with summable errors still converges weakly to solutions of the inclusion problem, if any.

In this section we consider the inclusion problem

$$0 \in A(x) + N_C(x)$$

where

k1) $A : X \to X$ is monotone and L-Lipschitz continuous ($L > 0$);
k2) N_C is the normal cone operator of $C \subset X$, a non-empty closed convex set.

KORPELEVICH'S METHOD
Choose $0 < \underline{\lambda} \leq \bar{\lambda} < 1/L$ and $x_0 \in X$;
for $k = 1, 2, \ldots$
a) choose $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$ and define

$$y_k = P_C(x_{k-1} - \lambda_k F(x_{k-1})), \qquad x_k = P_C(x_{k-1} - \lambda_k F(y_k)), \tag{18}$$

where P_C stands for the orthogonal projection onto C.

In order to cast this method in the formalism of Section 3, define for $\lambda > 0$

$$H_\lambda : X \to X \times X, \qquad H_\lambda(x) = (x, P_C(x - \lambda A(x))), \tag{19}$$

$$G_\lambda : X \times X \to X, \qquad G_\lambda(x, y) = P_C(x - \lambda A(y)). \tag{20}$$

Observe that since $P_C = J_{\lambda N_C}$, the second component of the (generic) operator H_λ is $J_{\lambda B} \circ (I - \lambda A)$ with $B = N_C$ which is the generic iteration map of the forward backward method in (10) (with $B = N_C$). Note also that the map H_λ above defined has an equivalent expression

$$H_\lambda(x) = (x, J_{\lambda N_C}(x - \lambda A(x)))$$

which can be obtained by setting $B = N_C$ in (13) $B = N_C$. Trivially,

$$x_k = G_{\lambda_k} \circ H_{\lambda_k}(x_{k-1}), \qquad k = 1, 2, \dots \tag{21}$$

The next two result were essentially proved in [14], in the context of the Hybrid Proximal-Extragradient Method. We will state and prove them in the context of σ-approximate resolvents.

Lemma 7.1. *If A and C satisfy assumptions* **k1** *and* **k2**, *then, for any $\lambda > 0$ and $x \in X$*

$$G_\lambda \circ H_\lambda(x) \in J_{A+N_C,\sigma}(x)$$

for $\sigma = \lambda L$.

Proof. Take $x \in X$ and let

$$y = P_C(x - \lambda A(x)), \quad z = P_C(x - \lambda A(y)).$$

Note that $z = G_\lambda \circ H_\lambda(x)$. Define

$$\begin{aligned}
\eta &= \frac{1}{\lambda}(x - \lambda A(x) - y),\\
\nu &= \frac{1}{\lambda}(x - \lambda A(y) - z), \quad \varepsilon = \langle \nu, z - y\rangle, \quad v = \nu + A(y).
\end{aligned}$$

Trivially, $\eta \in N_C(y)$ and $\nu \in N_C(z) = \partial\delta_C(z)$. Therefore,

$$\nu \in \partial_\varepsilon \delta_C(y) \subset (\partial\delta_C)^{[\varepsilon]}(y) = (N_C)^{[\varepsilon]}(y)$$

and

$$v \in (A + N_C)^{[\varepsilon]}(y), \qquad z = x - \lambda v. \tag{22}$$

Therefore

$$\begin{aligned}
\|\lambda v + y - x\|^2 + 2\lambda\varepsilon &= \|y - z\|^2 + 2\lambda\langle \nu, z - y\rangle\\
&= \|y - z\|^2 + 2\lambda\langle \nu - \eta, z - y\rangle + 2\lambda\langle \eta, z - y\rangle\\
&\le \|y - z\|^2 + 2\lambda\langle \nu - \eta, z - y\rangle,
\end{aligned}$$

where the inequality follows from the inclusions $\eta \in N_C(y)$, $z \in C$. Direct algebraic manipulations yield

$$\begin{aligned}
\|y - z\|^2 + 2\lambda\langle \nu - \eta, z - y\rangle &= \|\lambda(\nu - \eta) + z - y\|^2 - \|\lambda(\nu - \eta)\|^2\\
&\le \|\lambda(\nu - \eta) + z - y\|^2\\
&= \|\lambda(A(x) - A(y))\|^2.
\end{aligned}$$

Combining the two above equations, and using assumption **k1**, we conclude that

$$\|\lambda v + y - x\|^2 + 2\lambda\varepsilon \le (\lambda L)^2\|y - x\|^2.$$

The conclusion follows combining this inequality with (22). ☐

Corollary 7.2. *Let A, C be as in* **k1**, **k2**, *and $0 < \underline{\lambda} < \bar{\lambda} < 2\alpha$ and (λ_k), (x_k) be as in Korpelevich's method. Define*

$$\sigma = \bar{\lambda} L.$$

Then $0 < \sigma < 1$ and for any $x \in X$

$$G_{\lambda_k} \circ H_{\lambda_k}(x) \in J_{\lambda_k(A+B),\sigma}(x), \qquad k = 1, 2, \ldots$$

In particular

$$x_k = G_{\lambda_k} \circ H_{\lambda_k}(x_{k-1}) \in J_{\lambda_k(A+B),\sigma}(x_{k-1}), \qquad k = 1, 2, \ldots$$

and the sequence of maps $(G_{\lambda_k} \circ H_{\lambda_k})$ satisfies properties **P1**, **P2** *with respect to $(A + N_C)^{-1}(0)$.*

Proof. Use Lemma 7.1 and the same reasoning as in corollaries 5.2 and 6.2. □

Endowing $X \times X$ with the canonical inner product of Hilbert space products

$$\langle (x, y), (x', y') \rangle = \langle x, x' \rangle + \langle y, y' \rangle,$$

it is trivial to check that for $0 < \lambda \le \bar{\lambda}$, the maps H_λ and G_λ are Lipschitz continuous with constants

$$2 + \bar{\lambda} L, \qquad 1 + \bar{\lambda} L,$$

respectively. Hence, one can analyze Korpelevich's method with (summable) errors in the projections and/or evaluations of A etc.

8 Discussion

We provided a general definition of generic methods by means of recursive inclusions and sequences of point-to-set maps. Using this formulation, we defined two properties of those maps which guarantee that the associated method is Fejér convergent and generates sequences which converge to a solution, if any, even when perturbed by summable errors.

We think these results obviate the summable error convergence analysis of a number of convergent Fejér methods.

The framework for the analysis of Fejér convergent methods introduced here is, of course, not general enough to encompasses *all* of these methods. Indeed, if $X = \mathbb{R}$, $\Omega = \{0\}$ and

$$F(x) = \begin{cases} -x & x > 0, \\ x/2, & x \le 0 \end{cases}$$

then any sequence (x_n) satisfying $x_n = F(x_{n-1})$ is Fejér convergent to $\{0\}$ and converges to 0. However, the sequence $(F_n = F)$ does not satisfies **P2**.

It has been since long recognized that Korpelevich's method (and may be even the Forward-Backward method) was an "inexact" version of the proximal point method. However, the nature and degree of this "inexactness" were not known. We provided a formal definition of approximate solutions of the prox by means of the σ-approximate resolvent which, while encompassing many classical decomposition schemes, also guarantees weak convergence of sequences generated by such approximate resolvents (even in the presence of additional summable errors).

References

[1] Y. I. Alber, A. N. Iusem, and M. V. Solodov. Minimization of nonsmooth convex functionals in Banach spaces. *J. Convex Anal.*, 4(2):235–255, 1997.

[2] Y. I. Alber, A. N. Iusem, and M. V. Solodov. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Math. Programming*, 81(1, Ser. A):23–35, 1998.

[3] L. M. Briceño-Arias and P. L. Combettes. A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Optim.*, 21(4):1230–1250, 2011.

[4] R. Burachik, L. M. G. Drummond, A. N. Iusem, and B. F. Svaiter. Full convergence of the steepest descent method with inexact line searches. *Optimization*, 32(2):137–146, 1995.

[5] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, 5(2):159–180, 1997.

[6] R. S. Burachik, S. Scheimberg, and B. F. Svaiter. Robustness of the hybrid extragradient proximal-point algorithm. *J. Optim. Theory Appl.*, 111(1):117–136, 2001.

[7] R. S. Burachik and B. F. Svaiter. ϵ-enlargements of maximal monotone operators in Banach spaces. *Set-Valued Anal.*, 7(2):117–132, 1999.

[8] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504, 2004.

[9] J. M. Ermol′ev. The method of generalized stochastic gradients and stochastic quasi-Fejér sequences. *Kibernetika (Kiev)*, (2):73–83, 1969.

[10] J. M. Ermol′ev and A. D. Tuniev. Random Fejér and quasi-Fejér sequences. In *Theory of Optimal Solutions (Proc. Sem., Kiev, 1968), No. 2 (Russian)*, pages 76–83. Akad. Nauk Ukrain. SSR, Kiev, 1968.

[11] J. M. Ermol′ev and A. D. Tuniev. Random Fejér and quasi-Fejér sequences. In *Selected translations in mathematical statistics and probability. Vol. 13*, pages v+298. American Mathematical Society, Providence, R.I., 1973.

[12] A. N. Iusem, B. F. Svaiter, and M. Teboulle. Entropy-like proximal methods in convex programming. Research-Report 92-09, Mathematics University of Maryland, May 1992.

[13] A. N. Iusem, B. F. Svaiter, and M. Teboulle. Entropy-like proximal methods in convex programming. *Math. Oper. Res.*, 19(4):790–814, 1994.

[14] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim.*, 20(6):2755–2787, 2010.

[15] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.

[16] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[17] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[18] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[19] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.

[20] B. F. Svaiter. A family of enlargements of maximal monotone operators. *Set-Valued Anal.*, 8(4):311–328, 2000.

[21] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446, 2000.